Niveau(x) d'enseignement: S Numéro EDUQ : 8270719
Centre d'intérêt: 751

Bureau, Conrad, 1939-.

Rapport d'enquête sur la qualité du français écrit des étudiants de la Faculté des lettres de l'Université Laval / par Conrad Bureau avec la collab. de Benoît Papillon. -- [Québec] : Université Laval, Faculté des lettres, 1976.
90 f. ; 28 cm.

Cote Dewey: 440.9714/B952r Cote LC: PC/3608

Descripteur(s): *Français (Langue) *Qualité de la langue *Langue écrite *Grammaire *Orthographe *Compétence de l'étudiant *Universitaire (Niveau)

L'étude vise à connaître la langue écrite effectivement utilisée par les étudiants de l'Université Laval. Il s'agit d'obtenir, aux fins d'analyse, un corpus reflétant un état de la langue. L'expression "qualité du français écrit" renvoie dans ce contexte à un emploi de la langue conforme aux conventions linguistiques du français écrit.

L'enquête a eu lieu en mars 1975 à l'intérieur de 26 cours de la Faculté des lettres de l'Université Laval. On a demandé aux étudiants de ces cours de remplir une fiche de renseignements (âge, sexe, langue, profession du père, etc.) et de rédiger un court texte (une vingtaine de phrases) sur un sujet de leur choix. Sur un total possible de 1173 étudiants, 570 ont répondu soit 48,5%. Une triple correction de ces textes a permis de regrouper les types d'erreur selon les catégories suivantes: grammaire, orthographe, sémantique, syntaxe, anglicisme, renvoi ambigu.

En moyenne, les correcteurs ont relevé deux à trois fautes par page. Les fautes de grammaire comptent pour 33,8% des erreurs rencontrées, les fautes d'orthographe pour 26,2%, celles de syntaxe pour 17,2% et les anglicismes pour 3,9%. La moyenne cumulative de l'étudiant et le fait d'avoir étudié dans un établissement public ou privé n'ont pas d'incidence significative sur les résultats. Par contre, le sexe et le lieu d'origine, tout comme le niveau d'études et le programme d'études, semblent avoir un impact sur la qualité du français écrit.

Type de document: r
Région(s) administrative(s): 03
Auteur de l'analyse: ar

E D U Q 8270719

EDUQ

RAPPORT D'ENQUÊTE
SUR
LA QUALITÉ DU FRANCAIS ÉCRIT
DES ÉTUDIANTS DE LA FACULTÉ DES LETTRES
DE L'UNIVERSITÉ LAVAL

Université Laval, Faculté des lettres, février 1976

Copie pour EDUQ
Conrad Bu---
Le 15.09.82.
EDUQ 8270719

RAPPORT D'ENQUÊTE

sur

LA QUALITÉ DU FRANÇAIS ÉCRIT des ÉTUDIANTS de la FACULTÉ des LETTRES de l'UNIVERSITÉ LAVAL.

par

CONRAD BUREAU

Professeur adjoint,
Département de langues et linguistique,
Université Laval

avec la collaboration de

BENOÎT PAPILLON

Etudiant,

Département d'économique

Université Laval, 19 février 1976.

Nous tenons à remercier les étudiants
et les professeurs qui ont accepté de
collaborer à cette enquête,
les membres de la C.E.F.L. et tout
spécialement le Secrétaire de la
Faculté des lettres, M. Antonio Piché,
Madame Thérèse Lambert et Mademoiselle
Louise Rochette, toutes deux secrétaires
à la Faculté des lettres, pour leur
assistance technique,
Madame Michèle Rémillard, Professeur
assistant au Département de langues et
linguistique (Traduction) et Mademoiselle
Diane Potvin (Etudiante) pour la correction
des copies,
et Monsieur Benoît Papillon (Etudiant)
qui s'est occupé de l'aspect statistique
de cette enquête.
Merci également à madame Danielle Tessier
qui a dactylographié ce rapport.

"Feindre d'utiliser des
concepts rigoureux en les
dépouillant de leur rigueur
est une imposture."
(Gilles Granger)

PREMIÈRE PARTIE

L'ENQUÊTE

I INTRODUCTION

Origine de l'enquête

L'idée de mener une enquête sur la qualité du français écrit des étudiants n'est pas le fait de l'auteur du présent rapport. Ce sont les membres (1) de la Commission des études de la Faculté des lettres (C.E.F.L.) qui ont été les premiers à s'intéresser officiellement à cette question, comme en font foi les procès-verbaux.

A la suite de plusieurs discussions concernant "le français des étudiants", les membres de cette Commission ont conclu qu'il fallait recueillir des données sur le sujet; restait à déterminer la façon de procéder. C'est à ce moment que l'auteur du présent rapport a offert sa collaboration... bien imprudemment peut-être, car on ne pouvait prévoir alors tout le temps et tout le travail qui seraient nécessaires à la réalisation de cette entreprise. Cela explique, du moins en partie, que notre rapport d'enquête soit déposé plus tard qu'on ne l'avait envisagé au début.

---

1) On sait que la C.E.F.L. est composée d'un Vice-doyen, du Secrétaire de la Faculté et des Directeurs d'études des secteurs suivants: Arts et civilisations, Etudes françaises, Géographie, Histoire, Langues et littératures modernes et Langues secondes.

Il faut souligner ici un point important: le responsable de l'enquête a été mandaté pour organiser l'enquête, effectuer le dépouillement des données et proposer une analyse statistique — et c'était là son seul mandat. La nature de la tâche qui nous a été confiée nous amène à faire quelques considérations sur le point de vue du linguiste face à une enquête de ce type.

II QUELQUES ASPECTS THÉORIQUES

Linguistique et purisme

La linguistique est ou se veut science du langage et des langues; elle ne peut prétendre être telle si elle ne procède pas objectivement, ce qui implique que le linguiste doit s'interdire, en cours d'analyse, tout jugement de valeur. Bien sûr, on a beau pratiquer la linguistique, on n'en demeure pas moins un homme, un citoyen et quelqu'un dont la langue maternelle est, par exemple, le français; et à ce titre, on peut sans doute se permettre de porter un jugement sur la façon de parler ou d'écrire d'une personne ou d'un groupe. Mais dès qu'il s'agit de conduire une recherche linguistique, le linguiste, en tant que linguiste, doit se garder d'une telle attitude, sans quoi le terme science appliqué à la linguistique ne renvoie plus qu'à une prétention sans fondement dans la réalité.

C'est pourquoi notre tâche s'est limitée à l'observation, à la description et à l'analyse des faits tels qu'ils se sont présentés. Qu'on ne se surprenne donc pas si notre rapport refuse de prendre pour base d'analyse des notions comme "bon français", "mauvais français", "joual", "langue détériorée" et autres... autant de termes, d'expressions et de qualificatifs qui pleuvent dès qu'il s'agit de "juger" – mais non de décrire – la langue d'ici!

Mais la présence du terme "qualité", dans le nom même qui a été donné à cette enquête, ne va-t-elle pas à l'encontre du principe que nous venons d'énoncer?

Définitions

Nous entendons par qualité du français écrit: un emploi conforme aux conventions linguistiques qui régissent le code du français écrit. Il faut entendre par conventions linguistiques non pas un ensemble de règles imposées par une "élite" ou dictées par le "génie" (?) de la langue, mais bien les conventions de fait qui constituent en quelque sorte, par rapport au sujet parlant, le passé de la langue et qui sont nécessaires pour assurer la communication et/ou l'expression dans cette langue. C'est une convention du français que de renvoyer à la réalité bois par l'ensemble de

phonèmes /bwa/ ou par la graphie bois plutôt que par celle de bios , de tree ou de skov , et c'en est une autre d'écrire il dira et non il diront . Ces conventions ne sont pas d'ordre transcendant ni issues d'une sorte de totalitarisme linguistique: elles assurent simplement le fonctionnement d'une langue selon les règles d'organisation propres à cette langue.

Ainsi définie, la notion de "qualité" du français prend une valeur strictement technique et s'avère opératoire pour le type de recherche dont le présent rapport fait état.

Malgré tout, l'utilisation du terme "qualité" sera sans doute critiquée par certains, comme cela est arrivé quand l'enquête a eu lieu. Signalons à ce sujet que des chercheurs de l'Université de Montréal ont fait une enquête analogue à la nôtre auprès des cégépiens du Québec: il s'agit de l'Enquête sur le français écrit au collégial (ÉFEC). Nous aurions pu également dire "état" au lieu de "qualité", mais quel que soit le terme choisi, on peut toujours s'attendre à être critiqué, car en cette matière, on ne peut satisfaire tout le monde et son père... Nous pensons que ce qui importe, c'est moins le terme en soi que le fait d'en proposer une définition afin d'indiquer clairement dans quelle perspective on se situe. L'honnêteté du chercheur implique qu'il définisse

les termes de base de sa recherche et selon le point de vue de sa discipline. D'autres définitions sont certes possibles, mais il aurait été plus utile qu'elles nous fussent communiquées par les intéressés au moment de l'enquête ou du dépouillement des données et non après...

Ajoutons que si nous avons retenu le terme "qualité", c'était également pour des raisons de commodité: nous voulions en effet continuer à employer un terme qui était déjà utilisé par les membres de la C.E.F.L. et qui leur était familier. Enfin, comme nous avons pu le vérifier au cours de l'enquête, les étudiants ont très bien compris ce qu'il fallait entendre par "qualité"; c'est là un argument important en faveur du choix de ce terme, car c'est avant tout aux individus visés par une enquête que l'objet de cette enquête doit apparaître clairement.

Ceci dit, précisons le sens de ce terme dans le cadre de notre recherche. Puisque, par le terme "qualité" nous renvoyons strictement à la conformité avec le code, cela signifie que l'aspect stylistique par exemple ne fait pas partie de l'objet de notre investigation; loin d'être inscrit dans le code, le style est en effet un "surcodage", comme l'a bien montré Gilles-G. Granger dans

son Essai d'une philosophie du style. La définition que nous utilisons implique encore que nous n'avons pas à mesurer l'étendue du vocabulaire des sujets, la conformité au code n'étant pas nécessairement liée à la possession d'un vaste bagage lexical.

En résumé, nous avons pris, comme point de référence, le code effectif du français écrit, et non un quelconque modèle du "beau" langage imposé comme norme et qui serait "la meilleu façon d'écrire des meilleurs écrivains", ou encore "le meilleur français de la meilleure partie de la société"...parisienne ou autre... ou enfin "le français des meilleurs rédacteurs et annonceurs de Radio-Canada"! Tous ces superlatifs, on le voit, reflètent des jugements de valeur explicables peut-être mais inacceptables dans le cadre d'une recherche objective.

En linguistique, on ne peut pas considérer comme ayant une valeur scientifique toutes ces "normes" qui sont fondées sur le sentiment de chacun; seule une norme d'ordre statistique est valable, encore qu'on peut se demander si le terme "norme" convient en pareil cas. Or, il n'existe pas d'autre étude sur la qualité du français au niveau universitaire qui aurait pu nous servir, sinon de norme, du moins de base de comparaison. Nous considérons

donc le présent rapport comme la description d'un phénomène linguistique réel et bien délimité, description qui pourrait par la suite servir de point de comparaison à d'autres enquêtes de ce type.

III LES MODALITÉS DE L'ENQUÊTE

Test ou texte?

Dans le cas d'une enquête comme celle dont il est question ici, il y a deux façons d'obtenir les données: on peut fabriquer un test et le faire subir à la population visée, ou encore demander aux individus de produire un texte sur un sujet de leur choix.

Un test a les avantages suivants: uniformité, précision des questions, facilité de correction. Mais il a l'inconvénient d'être directif; en effet quelles que soient les précautions que l'on prenne et même si l'on va jusqu'à procéder à une validation du test, il n'en reste pas moins qu'avec une telle formule, les auteurs du test doivent faire un choix parmi les questions possibles, de telle sorte que les réponses des sujets soumis au test sont, par définition, "orientées", Le test nous semble très valable pour vérifier des connaissances, mais beaucoup moins lorsqu'il

s'agit d'obtenir un corpus reflétant "un état de langue". En somme, le but de l'enquête n'étant pas de vérifier les connaissances des étudiants dans le domaine de leur langue d'usage, ni de leur "poser des colles", mais bien d'étudier la langue écrite qu'ils emploient effectivement, il nous fallait obtenir un échantillon, un spécimen le plus naturel possible de la langue de chaque sujet soumis à l'enquête; c'est pourquoi nous avons opté pour la formule de la rédaction d'un texte sur un sujet libre.

La seule contrainte imposée aux étudiants est qu'ils devaient rédiger leur texte sur commande et peut-être à un moment où ils n'avaient pas le goût de le faire; mais les conditions psychologiques auraient été les mêmes si nous avions administré un test. Nous avons toutefois signalé aux étudiants que cette enquête n'avait rien d'obligatoire et qu'ils pouvaient, s'ils le désiraient, refuser de participer: plusieurs d'entre eux se sont d'ailleurs prévalus de ce droit.

En plus de rédiger un texte, les étudiants qui acceptaient de se soumettre à l'enquête devaient remplir une fiche de renseignements dont nous avions besoin pour les fins de l'analyse statistique.

La fiche de renseignements

Une des règles de notre enquête étant de respecter l'anonymat des étudiants, ces derniers n'avaient pas à écrire leur nom sur la fiche; cependant, comme nous leur demandions de donner plusieurs renseignements les concernant, certains étudiants ont vu là un moyen détourné de les identifier,... comme si des chercheurs avaient le temps de s'amuser à jouer au Sherlock Holmes de la langue. L'éthique professionnelle, pour ceux qui l'ignoreraient, cela existe aussi chez les linguistes... surtout dans le cas d'une enquête de ce type.

Voici donc la liste des renseignements demandés aux participants:

Niveau scolaire: _____

Programme d'études (diplôme postulé) _____

Âge _____

Sexe _____

Langue maternelle _____

Langue usuelle _____

Lieu de naissance _____
Ville _____
ou
Campagne _____

Année de naissance _____

Résidence permanente (lieu) _____
Ville _____
ou
Campagne _____

Profession du père _____

Occupation du candidat _____

Moyenne cumulative _____

Dernière institution pré-universitaire fréquentée _____

S'agissait-il d'une institution privée ou publique _____

Ces renseignements constituent, comme on le voit, un ensemble de paramètres non linguistiques qu'il sera intéressant de mettre en relation avec les variables linguistiques au moment de l'analyse statistique.

A cette fiche, nous avons joint des feuilles (format 8½ x 14) sur lesquelles les étudiants devaient rédiger leur texte.

Le déroulement de l'enquête

Cette enquête n'aurait pas été possible sans la collaboration de plusieurs professeurs de la Faculté des lettres; il nous fallait en effet obtenir une heure d'un de leurs cours pour permettre aux étudiants de rédiger leur texte. Les responsables des

Directions d'études concernées ont déterminé des cours-cibles (35 au total), pour chaque secteur: Arts et civilisations (AC), Etudes françaises (EF), Histoire (HIS), Géographie (GEO), Langues et littératures modernes (LLM). Dans la plupart des cas, ces cours-cibles correspondaient à des cours obligatoires pour les étudiants du premier cycle d'études; c'était la meilleure façon d'atteindre tous les étudiants inscrits à un programme et d'éviter qu'ils soient soumis à l'enquête plus d'une fois. En pratique, un même étudiant pouvait se retrouver dans deux cours-cibles; cela s'explique par la possibilité qu'ont les étudiants de choisir des cours en dehors du programme-type de chaque champ d'études. Dans un tel cas, évidemment, l'étudiant ne rédigeait pas un nouveau texte; ce fait aura d'ailleurs une incidence sur le taux réel de participation par rapport à la population visée.

Nous avons atteint les étudiants de vingt-six (26) cours différents. Comme cette enquête s'est déroulée pendant une courte période de temps (du 10 au 14 mars 1975), il a été impossible à celui qui la dirigeait de se présenter devant chaque groupe. Dans la plupart des cas, ce sont les professeurs qui ont distribué les formulaires à leurs étudiants; dans les autres cas, ceux où les professeurs refusaient de superviser la réalisation matérielle de l'enquête dans leur classe, c'est soit le responsable, soit un(e)

assistant(e) qui le remplaçait.

Afin d'assurer une présentation uniforme, un texte (avec une coquille d'ailleurs!) expliquant la nature et les objectifs de l'enquête a été remis à chaque étudiant; bien qu'il y ait eu diverses personnes à se présenter devant des groupes différents, on peut donc dire que l'enquête a été réalisée dans les mêmes conditions pour chaque groupe.

Les étudiants avaient une heure pour remplir la fiche de renseignements et rédiger leur texte; en ce qui concerne la longueur de ce dernier, aucun maximum ni minimum n'a été imposé ni même suggéré.

La participation à l'enquête

D'après les chiffres fournis par le secrétariat de la Faculté, nous pouvions espérer atteindre 1,173 individus, par le biais des vingt-six cours-cibles; nous avons obtenu 570 réponses, ce qui représente près de la moitié (48.5%) des étudiants inscrits à ces cours. En calculant une moyenne de vingt (20) phrases par texte, nous obtenons un corpus d'environ 10,000 phrases — nombre qui nous semble amplement suffisant lorsqu'il s'agit de découvrir les habitudes linguistiques d'un groupe et donc de caractériser un état de langue.

Même si ce corpus est suffisant pour les besoins de l'analyse, on peut se demander pourquoi la participation n'a pas été plus forte? Des facteurs techniques d'une part et phychologiques d'autre part sont à l'origine de cette participation relative des étudiants.

L'une des causes de la différence entre le nombre réel de réponses et leur nombre possible, c'est la présence d'un même étudiant dans deux ou plus de deux cours définis comme cours-cibles. Parmi les autres causes, citons: les absences des étudiants et le fait que certains professeurs aient oublié, tout simplement, de distribuer les formulaires d'enquête. Ainsi, un certain nombre d'étudiants n'ont pu participer à cette enquête par suite de facteurs ou d'erreurs techniques dont nous n'étions pas responsables, étant donné le grand nombre de personnes appelées à collaborer à cette enquête.

A côté de ces facteurs techniques, il y a eu, croyons-nous, des facteurs d'ordre psychologique qui ont joué et qui peuvent rendre compte du reste des cas de non-participation. Les étudiants, répétons-le, avaient la liberté de participer ou de ne pas participer à cette enquête, et ils en étaient avertis. Plusieurs d'entre eux,

sans protester verbalement, ont refusé de se soumettre à l'enquête, en quittant simplement la salle de cours dès qu'ils avaient appris ce qu'on attendait d'eux. On peut penser que les raisons qui expliquent ou justifient une telle réaction sont les suivantes: indifférence, manque d'intérêt envers l'objet de l'enquête et ses résultats éventuels, crainte d'être pénalisé sournoisement et enfin l'attitude "ras-le-bol". En effet, si l'on tient compte du climat socio-politique et culturel qui règne au Québec depuis quelques années et particulièrement en ce qui a trait à la langue d'ici, il est fort compréhensible que, dans le contexte d'une Loi 22 et après tant de luttes, de débats, de colloques, d'articles et d'émissions radiophoniques ou télévisées autour du français, après tant de doléances surtout et tant de jugements dépréciatifs, les Québécois en général et les étudiants en particulier "en aient marre" de voir ce sujet presque quotidiennement à l'ordre du jour! De même qu'on parle en médecine de "sentiment de sympathie exacerbé", on peut définir cette attitude par l'expression: "conscience linguistique exacerbée". Nous pensons que c'est un état de conscience dont il faut tenir compte.

On pourrait nous reprocher de postuler ces différentes causes; il n'en est rien comme en font foi les opinions que plusieurs étudiants, parmi ceux mêmes qui ont répondu, ont décidé de nous

transmettre par écrit sur le sujet. Plus d'un en effet a profité de cette enquête pour dire ce qu'il pensait de la situation du français au Québec ou pour se plaindre du fait qu'on enquête une fois de plus à propos de la langue... en oubliant toutefois qu'on ne s'était guère intéressé jusqu'ici à l'aspect objectif et statistique de la question. D'autres étudiants, comme nous l'avons déjà mentionné, ont vu dans la fiche de renseignements un moyen déguisé de les identifier et peut-être de les pénaliser par la suite.

On peut donc affirmer, pour conclure, qu'il y a eu, jusqu'à un certain degré et comme il fallait s'y attendre, une sorte de boycottage tacite de l'enquête de la part des étudiants.

L'examen des copies

Rappelons que par qualité du français écrit, il faut entendre: un emploi conforme aux conventions linguistiques qui régissent le code du français écrit. Notre premier travail a donc consisté à identifier les cas de non-conformité par rapport à la grammaire, à l'orthographe, à la sémantique et à la syntaxe du français écrit; cette façon de procéder nous permettait de couvrir tous les cas. Nous avons toutefois établi deux autres catégories: renvois ambigüs et anglicismes. Tout anglicisme affecte évidemment soit la grammaire, soit l'orthographe, soit la syntaxe, soit la

sémantique, mais tout emploi non conforme au code, de quelque aspect qu'il s'agisse, n'est pas forcément un anglicisme. Nous avons donc cru bon de grouper tous les anglicismes dans une même classe. Quant à l'autre catégorie, considérons qu'il y a renvoi ambigü lorsque l'emploi d'un substitut (un pronom) est équivoque.

Voici des exemples de non-conformité pour chaque catégorie (voir la séquence soulignée):

- grammaire[1]: "Bien qu'il avait déjà tenté de présenter"... (copie no 15113, p.2)
- orthographe: "Si chacun si mettait"... (copie no 15145 p. 3)
- sémantique: "Mon sujet sera une réflexion sur les diverses enquêtes que nous devons fréquemment remplir". (copie no 15136, p. 1)
- syntaxe[1]: "La manière de vivre l'amour au temps de nos grand-parents (sic) est certes très différente de la manière que la génération actuelle le vit". (copie no 15132, p. 3)
- anglicisme: "Et il réalise d'autre part, que la connaissance de la sexualité est beaucoup plus évoluée dans les pays"... (copie no 15132, p. 3)
- renvoi ambigü: "Quelle est la réaction de l'étudiant qui se voit obligé de rédiger un texte de quelque longueur que ce soit, quand malheureusement celui-ci n'en a pas le goût?" (copie no 15162, p. 1)

---

[1] Pour les fins de ce rapport, la catégorie grammaire comprend tous les cas d'accord, de conjugaison et de concordance des temps, et la catégorie syntaxe, les cas relevant de la construction de la phrase.

Une fois ces catégories établies, nous avons ensuite procédé de la façon suivante: en premier lieu, le corpus a été partagé entre deux assistantes de recherche qui ont fait un premier examen des textes; les assistantes ont ensuite échangé les copies, de façon à pouvoir examiner l'autre partie du corpus et à vérifier en même temps l'exactitude du premier relevé; enfin, toutes les copies sans exception ont été revisées par l'auteur du présent rapport. Chaque copie a donc été l'objet d'un triple examen.

Malgré ces précautions, il y a sans doute des cas qui ont échappé à ces trois personnes, tout simplement parce qu'elles ne sont pas des machines... Toutefois, aurions-nous oublié une centaine de cas, que l'analyse statistique n'en serait pas affectée; on peut en effet estimer à environ 300,000 — et c'est là un chiffre conservateur — le nombre de cas à examiner et à vérifier: 10,000 phrases X 10 groupes fonctionnels en moyenne X 3 unités par groupe fonctionnel = 300,000 possibilités. Une centaine d'oublis de notre part équivaudrait à une marge d'erreur de 0.03 de 1%, ce qui est plus qu'acceptable.

Pour chaque copie, nous avons donc relevé tous les cas de non-conformité par rapport au code du français écrit, en inscrivant dans la marge un symbole identifiant la catégorie concernée:

A pour anglicisme, O pour orthographe, SE pour sémantique, ST pour syntaxe, G pour grammaire et RA pour renvoi ambigü. Afin de faciliter les vérifications, nous nous sommes imposés de récrire sur une feuille que nous avons ajoutée au dossier de l'étudiant, tous les cas de non-conformité observés dans son texte. Nous avons également calculé la longueur de chacun des textes, renseignement que nous avons inscrit sur la même feuille.

Toutes les données ainsi obtenues de même que celles fournies par la fiche de renseignements ont ensuite été codifiées et transcrites sur cartes perforées en vue du traitement automatique.

"Le signe survit à la chose signifiée." (Aragon)

DEUXIÈME PARTIE

ANALYSE STATISTIQUE

INTRODUCTION

Etant donné d'une part le nombre imposant de données dont nous disposons et d'autre part la nécessité de produire une analyse scientifique, nous ne pouvions nous contenter d'une étude superficielle fondée uniquement sur des nombres et des moyennes absolus. L'analyse statistique qui est à la source de cette deuxième partie du rapport a été très élaborée, au point que nous devrons ne considérer que les résultats qui nous apparaissent les plus pertinents, et ne reproduire que quelques-uns des centaines de tableaux que nous avons obtenus. L'ensemble des documents statistiques a été déposé au Secrétariat de la Faculté des lettres, pour fins de consultation (Dossier de l'enquête).

Pour traiter les réponses codées, nous avons utilisé la nouvelle version du groupe de programmes SPSS ("Statistical Package for the Social Sciences"), version disponible au Centre de traitement de l'information (C.T.I.) de l'Université Laval.

Le dossier de chaque étudiant nous fournissait, lorsqu'il était complet, les valeurs de huit paramètres: niveau, Direction d'études, âge, sexe, lieu d'origine, institution préuniversitaire fréquentée, moyenne cumulative et nombre de pages composées. Pour certaines copies, nous

ignorons la valeur de tel ou tel paramètre, l'étudiant n'ayant pas fourni le renseignement demandé, mais nous avons tenté d'exploiter au maximum l'information contenue dans les dossiers incomplets; la base de calcul varie donc pour plusieurs paramètres: elle est égale au nombre de données recueillies pour telle ou telle rubrique de la fiche de renseignements.

Pour tous les dossiers sans exception, nous connaissions les valeurs de six variables, soit le nombre de cas de non-conformité —ou de "fautes", si l'on préfère— dans les catégories grammaire, orthographe, sémantique et syntaxe, à quoi il faut ajouter le nombre de renvois ambigüs et le nombre d'anglicismes.

Le schéma I donne une classification des paramètres et des variables suivant leur type. Les paramètres nominaux sont des paramètres pour lesquels l'ordre ou la mesure ne sont pas pertinents; ainsi les chiffres de 1 à 5 qui identifient, dans le code, les différentes Directions d'études n'ont aucune pertinence statistique. Les paramètres ordinaux sont ceux où l'ordre et seulement l'ordre est pertinent; c'est le cas du niveau: deuxième année par rapport à première année. Enfin, on parle de paramètres ou de variables du type ratio, lorsque le rapport entre deux valeurs d'un même paramètre, d'une même variable ou de plusieurs a un sens; ainsi, le rapport entre le nombre de fautes de deux copies a un sens.

A partir de ces six variables, nous avons calculé d'autres variables, comme le nombre de fautes moyen de chaque catégorie, la proportion de fautes de chaque catégorie, le nombre total moyen de fautes par page composée, etc.

La première section de l'analyse statistique contient d'abord des considérations d'ordre théorique sur quelques aspects de la procédure d'enquête; on y traite surtout de recensement et d'échantillon, deux notions à clarifier étant donné que le nombre total de réponses reçues n'est pas égal au nombre de réponses qui étaient possibles; ce fait aura une incidence sur le mode d'analyse adopté. Nous donnons ensuite la répartition des individus interrogés suivant les valeurs de quelques-uns des paramètres.

La deuxième section présente des statistiques pour l'ensemble des individus qui ont répondu et pour chaque sous-groupe; on y trouvera aussi le résultat de nombreux tests concernant les différences entre les sous-groupes.

La troisième section, enfin, étudie les relations entre les variables.

Schéma I. Variables et paramètres*

Types	Noms
Paramètres nominaux	- Directions d'études - Sexe - Lieu d'origine - Institution préuniversitaire fréquentée
Paramètres ordinaux	- Niveau
Paramètres du type ratio	- Âge - Moyenne cumulative - Nombre de pages composées
Variables du type ratio	- Nombre de fautes de grammaire - Nombre de fautes d'orthographe - Nombre de fautes de sémantique - Nombre de fautes de syntaxe - Nombre d'anglicismes - Nombre de renvois ambigus

* Un élément peut être considéré comme un paramètre ou une variable, selon le type d'analyse envisagée.

I

L'ÉCHANTILLON. LA RÉPARTITION DES INDIVIDUS

Caractère de l'échantillon et limitations

Pour obtenir des statistiques sur une population à partir d'un échantillon, ce dernier doit être représentatif et avoir été tiré de façon aléatoire. Ces deux conditions ne sont pas parfaitement respectées ici. En effet, l'enquête s'adressait en principe à toute la population des étudiants inscrits aux différents cours-cibles; dans le cas où on rejoint tous les individus, on parle de recensement. Or, nous avons déjà établi qu'une partie seulement de la population visée a effectivement participé à l'enquête; il est inévitable que la participation soit relative quand les individus ont le choix entre le consentement ou le refus.

C'est donc à partir d'un échantillon plus ou moins aléatoire que nous devrons établir les résultats pour l'ensemble de la Faculté. En conséquence, la fiabilité des données qui seront présentées dans l'analyse qui va suivre pourrait être critiquée en invoquant tel ou tel argument relatif à la méthodologie de la statistique. Toutefois, compte tenu de l'importance de la taille de l'échantillon et de la simi-

litude (dans la mesure où il a été possible de la vérifier) entre d'une part la distribution à l'intérieur de l'échantillon réel, et d'autre part la distribution à l'intérieur de la population totale, ces statistiques sont des indicateurs valables de la qualité du français écrit des étudiants de la Faculté des lettres. En effet, comme on le verra par la suite, nous avons effectué, chaque fois que c'était nécessaire, des tests statistiques (calcul du rapport F et calcul du T) pour vérifier la fiabilité des résultats, c'est-à-dire pour savoir si la différence entre les moyennes des différents sous-groupes pouvait être attribuée à une erreur d'échantillonnage. Toutes les précautions ont donc été prises pour éviter de faire "parler" faussement les résultats.

Répartition des individus d'après l'un des paramètres

La population-cible était d'environ 1,200 étudiants; nous avons reçu 570 réponses dont 545 ont été retenues pour l'analyse statistique; les copies éliminées sont celles d'étudiants n'appartenant pas au premier cycle d'études ou dont la langue usuelle n'était pas le français.

Nous présentons, dans les tableaux 1 à 5, la répartition des individus en tenant compte d'un seul paramètre à la fois: niveau, Direction d'études, sexe, lieu d'origine, type d'institution fréquentée au collégial. Comme ces tableaux sont très faciles à comprendre et qu'ils ne demandent pas

d'interprétation[1], nous les proposons tels quels (voir pp. 30-31) aux lecteurs, en nous limitant aux commentaires suivants:

- Les étudiants de première et de deuxième année représentent chacun à peu près le tiers de l'effectif total (Tableau I);
- Les étudiants appartenant à la Direction d'études "Langues et littératures modernes" constituent le groupe le plus important de l'échantillon, soit 45.3% (Tableau 2).
- Notre échantillon est composé d'individus de sexe féminin dans une proportion de plus de 55% (Tableau 3).
- 63% des individus viennent de la ville (Tableau 4).
- Enfin, les étudiants ayant fréquenté une institution préuniversitaire publique représentent 70% de l'échantillon (Tableau 5).

---

1. La fréquence absolue renvoie au nombre total d'étudiants qui ont fourni le renseignement demandé, tandis que la fréquence relative représente le même groupe en pourcentage; inconnu signifie que le renseignement n'a pas été fourni par n étudiants. Nous n'avons pas cru nécessaire d'inclure dans les tableaux 1 à 5 la fréquence ajustée et la fréquence cumulative qui apparaissent pourtant sur les originaux.

Tableau 1. Répartition des individus d'après le niveau		
	Fréquence absolue	Fréquence relative
Première année	190	34.9
Deuxième année	183	33.6
Troisième année	96	17.6
Inconnu	76	13.9
Total	545	100.0

Tableau 2. Répartition des individus d'après la Direction d'études		
	Fréquence absolue	Fréquence relative
EF	103	18.9
AC	40	7.3
GEO	70	12.8
HIS	85	15.6
LLM	247	45.3

Tableau 3. Répartition des individus d'après le sexe

	Fréquence absolue	Fréquence relative
Masculin	238	43.7
Féminin	302	55.4
Inconnu	5	0.9

Tableau 4. Répartition des individus d'après le lieu d'origine

	Fréquence absolue	Fréquence relative
Ville	340	62.4
Campagne	198	36.3
Inconnu	7	1.3

Tableau 5. Répartition des individus d'après l'institution préuniversitaire fréquentée

	Fréquence absolue	Fréquence relative
Privée	150	27.5
Publique	383	70.3
Inconnu	12	2.2

Répartition des individus d'après deux paramètres

Les tableaux 6 à 14 établissent la répartition des étudiants en tenant compte de deux paramètres à la fois. Nous ferons ici quelques remarques, pour aider les lecteurs à lire ces tableaux comme il se doit. Les quatre nombres qui apparaissent dans chacun des rectangles du tableau renvoient, dans l'ordre, aux données suivantes: le nombre de répondants, le pourcentage pour la rangée, le pourcentage pour la colonne et le pourcentage par rapport à la population totale de l'échantillon. Le nombre total et le pourcentage total apparaissent à l'extrême droite pour chaque rangée, et au bas pour chaque colonne. Le symbole 0 (zéro) qu'on rencontre parfois au sommet d'une colonne signale toujours les cas où le renseignement n'a pas été fourni. Comme précédemment, nous allons d'abord faire quelques commentaires concernant les tableaux et présenter ensuite ces derniers aux pages 37-45.

Niveau et Direction d'études (Tableau 6)

Comme ce sont les étudiants de "langues et littératures modernes" (LLM) qui constituent le groupe le plus important de l'échantillon (45.3%), il est normal de retrouver une proportion plus grande d'individus de cette Direction d'études à chaque niveau: 42.6% des individus de première année, 44.8% de la population de deuxième année et 55.2% des étudiants de troisième année.

D'autre part, si l'on considère chaque Direction d'études en elle-même, le tableau 6 révèle que les étudiants de première et de deuxième année forment un groupe nettement majoritaire—ce à quoi il fallait s'attendre, le nombre d'étudiants ayant tendance à diminuer à mesure que les études avancent. Voici les pourcentages cumulatifs (étudiants de première et de deuxième années) pour chaque Direction d'études:

Etudes françaises: 64.1%
Arts et civilisations: 75.0%
Géographie: 72.9%
Histoire: 74.1%
Langues et littératures modernes: 66.0%

Direction d'études et sexe (Tableau 7)

Le groupe féminin est majoritaire dans trois Directions d'études: EF = 60.2%, AC = 70.0%, LLM = 63.6%, tandis que le groupe masculin domine en Géographie (64.3%) et en Histoire (62.4%). D'après ces données, le sexe aurait une incidence sur le choix du programme d'études. On sait d'autre part que le groupe féminin a participé à l'enquête dans une proportion de 55.4% de l'échantillon total, mais il faudrait se garder de conclure que le paramètre sexe a eu une influence sur la participation à l'enquête. En effet, la comparaison des données obtenues pour Histoire et Géographie d'une part, avec celles obtenues pour Études françaises, Arts et civilisations et Langues et littératures modernes d'autre part, démontrent

que si, globalement, le groupe féminin a davantage répondu, cela tient simplement au fait d'une représentation plus forte de ce groupe au sein des trois Directions d'études qui ont fourni le plus de participants à l'enquête: 71.5% des répondants appartiennent aux trois Directions d'études EF, AC et LLM comme le montre le tableau 2.

Direction d'études et institution préuniversitaire fréquentée (Tableau 8)

Nous ferons ici une seule observation: la grande majorité des étudiants ont fréquenté une institution publique avant d'entreprendre les études universitaires, et ce, quelle que soit la Direction d'études concernée: le pourcentage le plus bas est de 68.2% (HIS) et le plus élevé est de 74.3% (GEO).

Direction d'études et lieu d'origine (Tableau 9)

La répartition des étudiants dans les différentes Directions d'études selon qu'ils viennent de la ville ou de la campagne est très inégale. Ainsi, le groupe citadin, qui domine nettement dans la plupart des Directions d'études, se répartit comme suit: AC = 65.0%, GEO = 81.4%, HIS = 64.7%, LLM = 61.5%; ce même groupe représente 62.4% de la population totale de notre échantillon. Il n'y a qu'un cas où l'effectif soit plus important pour le paramètre campagne, soit en Etudes françaises (51.5%) —encore qu'il ne représente qu'un peu plus de la moitié de ce groupe de répondants.

Age et institution préuniversitaire fréquentée (Tableau 10)

En regroupant les étudiants en quatre classes d'âge, on remarque que la proportion d'individus ayant fréquenté une institution préuniversitaire privée augmente sensiblement pour les deux derniers groupes: 0-20 = 29.8%, 21-24 = 21.6%, 25-30 = 34%, 31 et plus = 50%. Ces résultats sont directement reliés au fait que l'implantation d'un réseau collégial public au Québec est relativement récente. Notons enfin que la classe 21/24 ans regroupe 51.7% de la population et que le quart des étudiants (25.9%) ont vingt ans ou moins.

Lieu d'origine et institution préuniversitaire fréquentée (Tableau 11)

Qu'ils viennent de la ville ou de la campagne, les étudiants de notre échantillon ont fréquenté une institution préuniversitaire publique dans une proportion de 70% environ.

Niveau et institution fréquentée (Tableau 12)

On observe la même importance de l'institution publique si on la met en rapport avec le paramètre niveau: 73.7% en première année, 69.4% en deuxième et 72.9% en troisième.

Niveau et âge (Tableau 13)

Puisque les étudiants de la classe 21-24 ans constituent plus de la moitié de la population, il est normal que cette classe fournisse les effectifs les plus importants

dans deux cas sur trois: appartiennent à cette classe d'âge 66.1% des étudiants de deuxième année et 70.8% des étudiants de troisième année.

D'autre part, près de la moitié (48.9%) des étudiants de première année ont vingt ans ou moins.

Niveau et sexe (Tableau 14)

La proportion d'individus de sexe féminin dans les différents niveaux est sensiblement la même que la proportion observée pour la population totale: 55.8% en première année, 54.6% en deuxième et 57.3% en troisième par rapport à 55.4% pour tout l'effectif.

TABLEAU 6*

Répartition des individus d'après le niveau et la Direction d'études						
	EF	AC	GEO	HIS	LLM	TOTAL
Première année	21 11.1 20.4 3.9	26 13.7 65.0 4.8	27 14.2 38.6 5.0	35 18.4 41.2 6.4	81 42.6 32.8 14.9	190 34.9
Deuxième année	45 24.6 43.7 8.3	4 2.2 10.0 0.7	24 13.1 34.3 4.4	28 15.3 32.9 5.1	82 44.8 33.2 15.0	183 33.6
Troisième année	25 26.0 24.3 4.6	3 3.1 7.5 0.6	9 9.4 12.9 1.7	6 6.3 7.1 1.1	53 55.2 21.5 9.7	96 17.6
Niveau inconnu	12 15.8 11.7 2.2	7 9.2 17.5 1.3	10 13.2 14.3 1.8	16 21.1 18.8 2.9	31 40.8 12.6 5.7	76 13.9
TOTAL	103 18.9	40 7.3	70 12.8	85 15.6	247 45.3	545 100.0

* Les quatre nombres de chaque rectangle renvoient, dans l'ordre aux données suivantes: le nombre de répondants, le % pour la rangée, le % pour la colonne et le % par rapport à la population totale de l'échantillon. Les nombres et % totaux apparaissent à l'extrême droite pour chaque rangée, et au bas pour chaque colonne.

TABLEAU 7

Répartition des individus d'après le sexe et la Direction d'études				
	0	M	F	TOTAL
EF	0 0.0 0.0 0.0	41 39.8 17.2 7.5	62 60.2 20.5 11.4	103 18.9
AC	0 0.0 0.0 0.0	12 30.0 5.0 2.2	28 70.0 9.3 5.1	40 7.3
GEO	0 0.0 0.0 0.0	45 64.3 18.9 8.3	25 35.7 8.3 4.6	70 12.8
HIS	2 2.4 40.0 0.4	53 62.4 22.3 9.7	30 35.3 9.9 5.5	85 15.6
LLM	3 1.2 60.0 0.6	87 35.2 36.6 16.0	157 63.6 52.0 28.8	247 45.3
TOTAL	5 0.9	238 43.7	302 55.4	545 100.0

TABLEAU 8

Répartition des individus d'après la Direction d'études et l'institution préuniversitaire fréquentée

	0	PRIVÉE	PUBLIQUE	TOTAL
EF	1 1.0 8.3 0.2	28 27.2 18.7 5.1	74 71.8 19.3 13.6	103 18.9
AC	0 0.0 0.0 0.0	11 27.5 7.3 2.0	29 72.5 7.6 5.3	40 7.3
GEO	0 0.0 0.0 0.0	18 25.7 12.0 3.3	52 74.3 13.6 9.5	70 12.8
HIS	4 4.7 33.3 0.7	23 27.1 15.3 4.2	58 68.2 15.1 10.6	85 15.6
LLM	7 2.8 58.3 1.3	70 28.3 46.7 12.8	170 68.8 44.4 31.2	247 45.3
TOTAL	12 2.2	150 27.5	383 70.3	545 100.0

TABLEAU 9

Répartition des individus d'après le lieu d'origine et la Direction d'études

	0	VILLE	CAMPAGNE	TOTAL
EF	0 0.0 0.0 0.0	50 48.5 14.7 9.2	53 51.5 26.8 9.7	103 18.9
AC	1 2.5 14.3 0.2	26 65.0 7.6 4.8	13 32.5 6.6 2.4	40 7.3
GEO	0 0.0 0.0 0.0	57 81.4 16.8 10.5	13 18.6 6.6 2.4	70 12.8
HIS	3 3.5 42.9 0.6	55 64.7 16.2 10.1	27 31.8 13.6 5.0	85 15.6
LLM	3 1.2 42.9 0.6	152 61.5 44.7 27.9	92 37.2 46.5 16.9	247 45.3
TOTAL	7 1.3	340 62.4	198 36.3	545 100.0

TABLEAU 10

Répartition des individus d'après l'âge et l'institution préuniversitaire fréquentée.

	0	PRIVÉE	PUBLIQUE	TOTAL
	6 60.0 50.0 1.1	2 20.0 1.3 0.4	2 20.0 0.5 0.4	10 1.8
0-20	2 1.4 16.7 0.4	42 29.8 28.0 7.7	97 68.8 25.3 17.8	141 25.9
21-24	2 0.7 16.7 0.4	61 21.6 40.7 11.2	219 77.7 57.2 40.2	282 51.7
25-30	0 0.0 0.0 0.0	24 34.3 16.0 4.4	46 65.7 12.0 8.4	70 12.8
31 +	2 4.8 16.7 0.4	21 50.0 14.0 3.9	19 45.2 5.0 3.5	42 7.7
TOTAL	12 2.2	150 27.5	383 70.3	545 100.0

TABLEAU 11

Répartition des individus d'après le lieu d'origine et l'institution préuniversitaire fréquentée

	0	PRIVÉE	PUBLIQUE	TOTAL
	6 85.7 50.0 1.1	0 0.0 0.0 0.0	1 14.3 0.3 0.2	7 1.3
VILLE	1 0.3 8.3 0.2	97 28.5 64.7 17.8	242 71.2 63.2 44.4	340 62.4
CAMPAGNE	5 2.5 41.7 0.9	53 26.8 35.3 9.7	140 70.7 36.6 25.7	198 36.3
TOTAL	12 2.2	150 27.5	383 70.3	545 100.0

TABLEAU 12

Répartition des individus d'après le niveau et l'institution préuniversitaire fréquentée

	0	PRIVÉE	PUBLIQUE	TOTAL
Première année	2 1.1 16.7 0.4	48 25.3 32.0 8.8	140 73.7 36.6 25.7	190 34.9
Deuxième année	0 0.0 0.0 0.0	56 30.6 37.3 10.3	127 69.4 33.2 23.3	183 33.6
Troisième année	0 0.0 0.0 0.0	26 27.1 17.3 4.8	70 72.9 18.3 12.8	96 17.6
Niveau inconnu	10 13.2 83.3 1.8	20 26.3 13.3 3.7	46 60.5 12.0 8.4	76 13.9
TOTAL	12 2.2	150 27.5	383 70.3	545 100.0

TABLEAU 13

Répartition des individus d'après le niveau et l'âge

	0	0-20	21-24	25-30	31 +	TOTAL
Première année	2 1.1 20.0 0.4	93 48.9 66.0 17.1	64 33.7 22.7 11.7	23 12.1 32.9 4.2	8 4.2 19.0 1.5	190 34.9
Deuxième année	1 0.5 10.0 0.2	30 16.4 21.3 5.5	121 66.1 42.9 22.2	17 9.3 24.3 3.1	14 7.7 33.3 2.6	183 33.6
Troisième année	0 0.0 0.0 0.0	4 4.2 2.8 0.7	68 70.8 24.1 12.5	17 17.7 24.3 3.1	7 7.3 16.7 1.3	96 17.6
Niveau inconnu	7 9.2 70.0 1.3	14 18.4 9.9 2.6	29 38.2 10.3 5.3	13 17.1 18.6 2.4	13 17.1 31.0 2.4	76 13.9
TOTAL	10 1.8	141 25.9	282 51.7	70 12.8	42 7.7	545 100.0

TABLEAU 14

Répartition des individus d'après le niveau et le sexe				
	0	MASCULIN	FÉMININ	TOTAL
Première année	1 0.5 20.0 0.2	83 43.7 34.9 15.2	106 55.8 35.1 19.4	190 34.9
Deuxième année	0 0.0 0.0 0.0	83 45.4 34.9 15.2	100 54.6 33.1 18.3	183 33.6
Troisième année	0 0.0 0.0 0.0	41 42.7 17.2 7.5	55 57.3 18.2 10.1	96 17.6
Niveau inconnu	4 5.3 80.0 0.7	31 40.8 13.0 5.7	41 53.9 13.6 7.5	76 13.9
TOTAL	5 0.9	238 43.7	302 55.4	545 100.0

II

VARIABLES ET PARAMÈTRES

Nombre de pages composées

Les étudiants qui ont participé à l'enquête ont produit des textes dont la longueur moyenne est de 2.056 pages; nous disposons donc d'un corpus de plus de 1,000 pages écrites. Le tableau 15 présente les données relatives à la variable nombre de pages composées (NPC).

Tableau 15. Nombre de pages composées (Variable NPC)

Moyenne	Ecart-type	Variance	Intervalle de variation	Minimum	Maximum
2.056	0.732	0.536	3.8	0.2	4.0

En utilisant cette dernière variable (NPC) et les six variables: nombre de fautes d'orthographe (MVAR1), nombre de fautes de grammaire (MVAR2), nombre de fautes de syntaxe (MVAR3), nombre de fautes de sémantique (MVAR4), nombre de renvois ambigüs (MVAR5) et nombre d'anglicismes (MVAR6), dont nous connaissions les valeurs pour tous les dossiers sans exception (545), treize variables ont été créées.

Nombres moyens de cas par page composée (Tableau 16)

Tous les individus n'ayant pas produit un texte de

même longueur, il était nécessaire de considérer, pour chacune des six catégories, le nombre moyen de fautes par page plutôt que le nombre absolu de fautes par dossier, afin de pouvoir établir des comparaisons et obtenir des moyennes significatives. En prenant comme base le dossier, nous n'aurions pu proposer aucune interprétation valable à partir des moyennes; ces dernières en effet n'auraient eu que peu de signification dans la mesure ou tel individu a écrit un texte d'une demi-page et tel autre un texte de quatre pages.

En parcourant le tableau 16, on remarque que c'est pour la grammaire et l'orthographe, dans l'ordre, que les moyennes sont les plus élevées; viennent ensuite les moyennes observées pour la syntaxe puis pour la sémantique. Enfin, aussi étonnant que cela puisse paraître, la moyenne d'anglicismes, qui n'est que de 0.097, se situe juste avant la moyenne de renvois ambigüs (0.027) qui est la plus basse —résultat qui contredit beaucoup d'affirmations gratuites à propos du "franglais" de nos étudiants. On pourrait faire la même remarque à propos de la syntaxe. Une conclusion importante en effet se dégage de la lecture de ce tableau: la moyenne de fautes par page est, pour la grammaire et l'orthographe, plus de deux fois supérieure à la moyenne de fautes de syntaxe et plus de huit fois supérieure à celle des anglicismes. Bien entendu, des tests qui porteraient sur la syntaxe et les anglicismes et qu'on ferait passer aux mêmes étudiants

pourraient conduire à des résultats fort différents et peut-être fort alarmants pour certains; mais dans ce domaine, les tests vérifient des connaissances sur un sujet, plutôt que de mesurer ou d'évaluer un échantillon de langue naturelle, laquelle est, par définition pour le linguiste, la langue!

Tableau 16. Nombre de cas de non-conformité pour chaque catégorie

Variables	Moyenne	Ecart Type	Variance	Intervalle de variation	Minimum	Maximum
MVAR1 Orthographe	0.852	1.392	1.938	15.0	0.0	15.0
MVAR2 Grammaire	0.897	1.195	1.427	10.0	0.0	10.0
MVAR3 Syntaxe	0.399	0.608	0.370	5.0	0.0	5.0
MVAR4 Sémantique	0.144	0.302	0.091	2.308	0.0	2.308
MVAR5 Renvois ambigüs	0.027	0.147	0.022	2.0	0.0	2.0
MVAR6 Anglicismes	0.097	0.322	0.104	5.0	0.0	5.0

Proportions pour chaque catégorie (Tableau 17)

Il est intéressant de connaître quelle est la proportion de cas de non-conformité de chaque catégorie par rapport au nombre total par page composée. Comme le révèle le tableau 17, les individus ont fait en moyenne entre deux et trois fautes (2.417) par page; signalons que le minimum a été de zéro (0) et le maximum de trente (30).

Les fautes de grammaire comptent, en moyenne, pour le tiers (33.8%) des fautes rencontrées et la proportion moyenne pour l'aspect orthographique est de 26.2%, tandis qu'elle est de 17.8% pour la syntaxe et de 3.9% seulement pour les anglicismes. Il va de soi que, dans ce tableau, les moyennes établissent la même hiérarchie que dans le précédent. Ces statistiques présenteront un intérêt particulier quand nous voudrons comparer l'échantillon avec différentes classes d'individus, comme la classe des étudiants d'un niveau donné ou d'une Direction d'études.

Tableau 17. Proportion de cas de chaque catégorie

VARIABLES	MOYENNE	ECART-TYPE	VARIANCE
P1 Orthographe	0.262	0.285	0.081
P2 Grammaire	0.338	0.318	0.101
P3 Syntaxe	0.178	0.262	0.069
P4 Sémantique	0.069	0.170	0.029
P5 Renvois ambigus	0.011	0.065	0.004
P6 Anglicismes	0.039	0.127	0.016
MTVAR Nombre total de fautes par page	2.417	2.611	6.818

Méthode d'analyse

Plusieurs approches étaient possibles pour déterminer l'influence de nos paramètres sur les variables. Pour éviter de transformer en variables discrètes les variables continues —avec tout l'arbitraire que cela comporte— nous avons opté pour la méthode qui consiste à diviser notre échantillon en sous-groupes, suivant les valeurs discrètes[1] des paramètres; cela fait, on compare ensuite les moyennes des variables des sous-groupes. On effectue cette comparaison en testant l'hypothèse d'égalité entre les moyennes.

Or, pour tester cette hypothèse, le calcul du khi carré (χ^2) n'est plus valable quand on procède par sous-groupe; c'est pourquoi nous avons utilisé le "rapport F"[2] pour effectuer nos tests. Pour vérifier l'hypothèse d'égalité entre les moyennes des divers sous-groupes suivant une division donnée de l'échantillon, on calcule le rapport F pour cette division —division par niveaux, par Directions d'études, etc. On compare ensuite cette valeur avec le F théorique[3], pour un niveau de confiance fixée à .05.

---

1 Si ce n'est pas le cas (pour l'âge par exemple), on effectue un recodage significatif dans la mesure du possible.

2 Par des calculs statistiques, on obtient des valeurs comme le khi^2, le rapport F, etc. dont l'utilisation dépend de la nature de l'étude.

3 Le F théorique est obtenu à partir d'une distribution théorique analogue à la distribution représentée par la courbe normale.

Si le rapport F calculé est inférieur au F théorique, nous accepterons l'hypothèse d'égalité et nous attribuerons les différences observées entre les moyennes des sous-groupes de l'échantillon à une erreur d'échantillonnage. Si le rapport F calculé est supérieur au F théorique, nous conclurons que les différences observées entre les moyennes des sous-groupes de notre échantillon sont trop importantes pour être justifiées par une erreur d'échantillonnage et, par conséquent, nous rejetterons l'hypothèse d'égalité entre les moyennes.

Dans le cas où notre échantillon se divise en deux sous-groupes pour un certain paramètre, dans le cas du sexe par exemple, un autre calcul statistique équivalent[1] sera utilisé pour tester l'hypothèse d'égalité, le calcul du T; encore là, un T calculé supérieur au T théorique conduira au rejet de l'hypothèse.

Tous ces tests ont pour fonction de vérifier si les différences sont significatives ou non, et il est absolument nécessaire de les effectuer, parce que dans un échantillon nous ne considérons qu'une partie de la population. Ces précautions ont été prises pour éviter l'arbitraire dans l'interprétation des données.

---

1 La statistique F pourra encore être utilisée, mais elle entraînera les mêmes conclusions puisque la distribution F(1,n) qui nous intéresse quand on a deux sous-groupes est une distribution T(n) élevée au carré.

Sept séries de tableaux nous ont donné les statistiques pour les sept paramètres étudiés: Direction d'études, sexe, lieu d'origine, institution préuniversitaire fréquentée, niveau, âge et moyenne cumulative. Comme nous avons obtenu plus de 100 tableaux statistiques pour étudier l'influence des paramètres sur les variables, il est impossible de reproduire ici cet ensemble et d'expliciter tous les détails qu'ils nous ont fourni. Les lecteurs désireux d'en savoir davantage pourront consulter ces tableaux qui ont été versés au "Dossier de l'enquête" (Secrétariat de la Faculté des lettres). Nous nous limiterons, dans ce rapport, aux principales données et nous indiquerons dans quels cas les différences entre les moyennes des sous-groupes sont significatives ou non; ce faisant, nous donnerons aux lecteurs les faits les plus pertinents et les plus importants; en effet, les nombreux tests qu'a entraînés l'étude de l'influence des paramètres sur les variables n'ont été effectués que pour déterminer ce qui est significatif et ce qui ne l'est pas.

Direction d'études

Le test F nous permet de rejeter l'hypothèse d'égalité entre les moyennes pour la variable nombre total de fautes par page (MTVAR). Si on considère les proportions des divers types de fautes, l'hypothèse d'égalité entre les moyennes

est rejetée pour l'orthographe et la grammaire. Et même si on accepte l'hypothèse d'égalité pour les autres types de fautes, il existe des différences marquées entre les Directions d'études. Voici quelques détails...

Les moyennes, pour le nombre total de fautes par page (MTVAR), sont les suivantes, par ordre décroissant:

AC = 3.549

GEO = 3.546

HIS = 3.256

EF = 1.923

LLM = 1.830

Fait intéressant, ce sont les groupes d'étudiants dont les programmes d'études sont orientés vers la langue (EF et LLM) qui affichent les moyennes les plus basses. Comme l'hypothèse d'égalité a été rejeté pour ces résultats, on peut donc conclure que le type de programme d'études a une incidence sur la qualité du français des étudiants. Cette conclusion se vérifie d'ailleurs si l'on considère les variables nombre de fautes de grammaire (MVAR2) et nombre de fautcs d'orthographe (MVAR1):

Orthographe		Grammaire	
HIS =	1.311	AC =	1.658
GEO =	1.229	GEO =	1.507
AC =	1.217	HIS =	1.037
LLM =	0.690	EF =	0.798
EF =	0.464	LLM =	0.594

On se rappellera que les cas de non-conformité au code pour les catégories grammaire et orthographe représentent une proportion de 60.0% du nombre total de fautes, que, pour ces mêmes catégories, les tests ont démontré que nos résultats ne pouvaient être attribués à un défaut de l'échantillonnage et qu'ils sont donc significatifs, au sens statistique du terme.

Pour la catégorie syntaxe, on observe les plus hautes moyennes en Histoire (0.628) et en Géographie (0.597), et la plus basse en Arts et civilisations (0.282), ce qui peut surprendre en regard des résultats que nous avons donnés plus haut; les moyennes sont également basses pour Etudes françaises (0.358) et Langues et littératures modernes (0.301); ce sont là des différences marquées entre les deux premières Directions d'études citées (HIS et GEO) mais non significatives d'après les tests, répétons-le.

Quant à la sémantique, on observe la moyenne la plus élevée pour le groupe Arts et civilisations; les différences entre les quatre autres groupes sont minimes; c'est le même groupe qui domine pour les anglicismes. Enfin, pour la catégorie renvois ambigüs, les valeurs sont très faibles pour tous les groupes de sorte qu'on ne peut se permettre aucune interprétation; notons seulement que c'est le groupe LLM qui a la moyenne la plus basse (0.015) et le groupe EF qui domine (0.058).

On peut dire que, globalement, c'est le groupe LLM qui commet le moins de fautes; comme ce groupe constitue la partie la plus importante de notre échantillon (45.3%), c'est un fait à signaler.

Sexe

L'hypothèse d'égalité entre les moyennes est rejetée par les tests, F ou T, pour la variable nombre total de fautes par page (MTVAR). Le groupe masculin fait plus de fautes (2.8093) que le groupe féminin (2.0415). Ceci est particulièrement vrai pour MVAR1, l'orthographe (M = 1.0019, F = 0.6976) et MVAR2, la grammaire (M = 1.1335, F = 0.7094). Les tableaux 18 et 19 présentent, pour ces deux variables, une partie de l'analyse de variance à laquelle nous avons procédé; c'est ce type d'analyse qui a été utilisé pour déterminer l'influence de chacun des six paramètres sur chacune des sept variables et qui nous a fourni plus d'une centaine de tableaux dont nous proposons aux lecteurs deux spécimens.

Quant aux autres types de fautes, bien que les différences soient peu marquées, on remarque que le groupe féminin domine pour les renvois ambigüs, mais que c'est encore le groupe masculin qui enregistre la plus haute moyenne dans la catégorie syntaxe.

Il n'existe pas de différences significatives entre les moyennes des deux groupes pour les variables proportion de fautes de... par page, sauf pour l'orthographe, catégorie ou le groupe masculin domine encore (M = 0.291, F = 0.242).

Notons que la valeur du paramètre sexe était inconnue pour cinq dossiers.

Tableau 18. Analyse de variance: Sexe-MVAR1 (orthographe)

Variables	Somme des moyennes	Moyenne	Ecart-type	Somme des carrés	N
M	238.449	1.002	1.311	407.544	238
F	210.662	0.698	1.197	431.300	302
Total	449.111	0.832	1.257	838.844	540

F = 7.9068 F théorique = 3.8

Tableau 19. Analyse de variance: Sexe-MVAR2 (grammaire)

Variables	Somme des moyennes	Moyenne	Ecart-type	Somme des carrés	N
M	269.772	1.133	1.408	469.812	238
F	214.226	0.709	0.934	262.478	302
Total	483.998	0.896	1.184	732.290	540

F = 17.5918 F théorique = 3.8

Lieu d'origine

Par campagne, il faut entendre, dans ce rapport, les villages et les petites localités éloignées des grands centres. Sept personnes n'ont pas fourni le renseignement demandé, soit moins de 2% de l'échantillon.

Quel que soit le test effectué, F ou T, il faut rejeter

l'hypothèse d'égalité entre les moyennes, pour la variable nombre total de fautes par page; le groupe campagne commet moins de fautes (2.057) que le groupe ville (2.545). Ce résultat se vérifie en particulier pour les catégories grammaire ($C = 0.738$, $V = 0.984$), orthographe ($C = 0.737$, $V = 0.871$) et syntaxe ($C = 0.335$, $V = 0.425$). Bien que les tests ne rejettent pas l'hypothèse d'égalité, soulignons que la moyenne d'anglicismes chez les citadins est supérieure à celle de l'autre groupe.

Pour les variables proportion de... par page, les tests ne révèlent pas de différences significatives entre les moyennes. Disons seulement que là où il y a le plus grand écart entre les deux groupes, c'est pour la catégorie sémantique, et que dans ce cas c'est le groupe campagne qui a la plus haute moyenne ($C = 0.082$, $V = 0.061$).

Âge

La valeur de ce paramètre est inconnue pour dix dossiers. Plusieurs regroupements par classes d'âge ont été étudiés; nous présentons dans ce rapport les résultats obtenus à partir de deux de ces regroupements.

Pour l'un des regroupements, nous avons déterminé quatre classes: 0-20 ans, 21-24 ans, 25-30 ans, 31 ans et plus. Si nous adoptons un niveau de confiance de $\alpha = 0.05$,

les tests révèlent des différences significatives entre les moyennes, pour les variables nombre total de fautes par page et nombre de fautes de grammaire par page. Pour un niveau de confiance de $\alpha = 0.01$, les tests ne révèlent plus de différences significatives. On ne peut donc rejeter l'hypothèse d'égalité entre les moyennes, dans le cas de ces deux variables, sans réticences.

Les valeurs des moyennes sont plus basses pour les étudiants de la classe 31 ans et plus que pour ceux des autres classes d'âge, et ce, pour la plupart des types de fautes: les individus du groupe le plus âgé, en effet, font beaucoup moins de fautes de grammaire (de deux à trois moins) et d'orthographe (la moitié moins) et emploient moins d'anglicismes (de deux à trois fois moins). Pour les variables proportion de ... par page, les valeurs calculées pour le rapport F sont toutes relativement faibles et le test n'est significatif dans aucun cas.

Nous avons, par ailleurs, fait un autre regroupement, en deux classes cette fois: 0-24 ans et 25 et plus. Aucun des tests n'indique de différence significative quelle que soit la variable considérée, si l'on prend un niveau de confiance de 0.01. Par contre, pour un niveau de confiance de 0.05, le T qui a été calculé à partir de deux variances différentes pour les deux groupes permet de conclure à l'existence d'une différence significative entre les moyennes observées pour la variable nombre de fautes de grammaire par page:

0-24 ans = 0.7142 et 25 ans et plus = 0.9369.

Les résultats de ces deux analyses ne permettent pas de conclure que la qualité du français s'améliore avec l'âge, mais seulement que les étudiants âgés de trente ans et plus font moins de fautes que les étudiants des autres sous-groupes; il n'y a pas amélioration progressive, puisqu'on n'observe pas de différences marquées entre les moyennes des divers sous-groupes de moins de trente ans.

Institution préuniversitaire fréquentée

Rappelons que les répondants ont été répartis en deux classes: institution privée et institution publique. Douze individus (2.2% de l'échantillon) n'ont pas fourni le renseignement demandé à cette rubrique.

Aucun des tests effectués ne révèle de différence significative entre les moyennes des deux groupes. En observant le nombre moyen de fautes de chaque catégorie par page composée, nous avons remarqué que les moyennes du groupe venant d'institutions privées sont légèrement inférieures à celles de l'autre groupe. Mais on ne peut tirer de cette observation aucune conclusion à l'effet que l'institution privée donnerait à ses étudiants une meilleure formation en français; d'abord les différences sont peu marquées et de plus elles n'apparaissent que si l'on considère les moyennes brutes, c'est-à-dire avant même d'avoir procédé aux tests de vérification. Or, les

vingt-sept tests effectués, aussi bien pour les variables nombre de... par page que pour les variables proportion de ... par page ont démontré que les différences entre les moyennes des deux groupes ne sont pas significatives.

Nous nous permettons de verser nos résultats au dossier de la querelle entre institutions privées et institutions publiques, —querelle à coup de statistiques (?) et dont les journaux ont fait état à l'automne.

Moyenne cumulative

Quatre-vingt-trois individus n'ont pas donné leur moyenne cumulative en remplissant le questionnaire, ce qui représente plus de quinze pour-cent de l'échantillon. Pour analyser l'influence du paramètre moyenne cumulative sur les variables, nous avons procédé à un double contrôle: nous avons calculé des coefficients de corrélation d'ordre zéro et des coefficients de corrélation partiels. Le coefficient de corrélation rend compte d'un certain parallélisme entre l'évolution de deux ou de plusieurs variables, mais ne témoigne pas nécessairement d'une relation de cause à effet.

L'étude des coefficients de corrélation d'ordre zéro ne fait apparaître aucune corrélation significative entre le paramètre moyenne cumulative et les sept variables nombre de ... par page (MTVAR, MVAR1, MVAR2, MVAR3, MVAR4, MVAR5, MVAR6).C'est pour les variables nombre total moyen de fautes

par page (MTVAR) et le nombre de fautes de sémantique par page (MVAR4) que les coefficients de corrélation avec la moyenne cumulative sont les plus élevés en valeur absolue: ils sont respectivement de -.064 et -.075.

De même, si nous passons à l'étude des coefficients de corrélation partiels, nous n'obtenons aucune corrélation significative pour les quinze combinaisons de paramètres ou de variables de contrôle[1] essayées. Aucun des coefficients calculés ne dépasse 0.08 en valeur absolue; on comprend, dès lors, pourquoi aucune corrélation n'est significative.

Tous les coefficients de corrélation obtenus, qu'ils soient d'ordre zéro ou partiels, sont négatifs. On pourrait donc supposer, si les tests étaient significatifs, que le nombre de fautes de chaque catégorie a une tendance à décroître à mesure que la moyenne cumulative augmente. Mais, comme nous l'avons déjà constaté, en aucun cas les tests n'ont été significatifs. On peut penser que l'impossibilité où nous sommes de proposer une interprétation à partir de tels résultats est peut-être due à l'approche adoptée, qui a été différente de celle utilisée pour l'étude de l'influence de tous les autres paramètres (calcul de coefficients de corrélation plutôt que du rapport F ou T).

---

1. Nous entendons par paramètres ou variables de contrôle, ceux ou celles dont on essaie d'annuler l'influence sur les variables ou paramètres pour lesquels on veut tester la corrélation. On le fait en supposant une relation linéaire entre les variables testées et la ou les variables de contrôle, ce qui permet de calculer de nouvelles variables (à tester) purifiées de l'influence des variables de contrôle (lire: variables ou paramètres dans cette dernière phrase).

Nous avons donc fait une seconde analyse, utilisant cette fois notre approche habituelle: nous avons comparé les moyennes pour les différentes variables à partir de deux types de regroupements.

Nous avons d'abord divisé les individus qui avaient fourni leur moyenne cumulative en deux classes: ceux dont la moyenne cumulative est supérieure à 4 d'une part, et d'autre part tous les autres. Pour aucune des variables, le test T ne révèle de différence significative entre les moyennes des deux groupes. On remarque seulement que les moyennes du premier groupe (MC = plus de 4) sont toujours légèrement inférieures à celles de deuxième groupe, sauf pour la syntaxe.

Si nous adoptons la division suivante: MC = 0-2.5, 2.5-3.5, 3.5-4.0, 4.0-5.0, les tests, encore là ne sont pas significatifs pour les différences entre les moyennes. Chaque groupe a fait en moyenne entre 2.0 et 2.5 fautes par page composée. Le deuxième groupe (2.5-3.5) enregistre les moyennes les plus élevées pour toutes les variables, excepté pour la catégorie sémantique (MVAR4).

Toutes ces données statistiques et tous ces tests nous amènent à la conclusion suivante: les individus qui ont une moyenne cumulative très forte (plus de 4) sont peut-être moins susceptibles que les autres de faire des fautes, de quelque catégorie qu'il s'agisse; mais il faut se rappeler, en acceptant une telle conclusion, qu'elle n'est justifiée formellement par aucun test statistique mais seulement par

le comportement des variables dans leur ensemble. Par ailleurs, l'étude des variables proportion de ... par page ne révèle rien de significatif.

Niveau

Au total, 469 étudiants (près de 80%) ont indiqué à quel niveau ils appartenaient.

Les tests nous donnent des F assez élevés pour les sept variables nombre de ... par page; seuls les variables MVAR4 (sémantique) et MVAR5 (renvois ambigüs) ont des F inférieurs à 2. Avec un niveau de confiance de 0.01, on rejette l'hypothèse d'égalité entre les moyennes dans le cas du nombre total de fautes (MTVAR) et de la catégorie syntaxe (MVAR3); avec un niveau de confiance fixé à 0.05, on rejette l'hypothèse d'égalité pour une autre variable: le nombre de fautes de grammaire (MVAR2). Dans tous ces cas où les différences sont significatives, la moyenne pour le troisième niveau est nettement inférieure à celle des deux premiers niveaux:

Variables	Première année	Deuxième année	Troisième année
MTVAR	2.666	2.405	1.634
MVAR3	0.446	0.440	0.242
MVAR2	1.017	0.895	0.621

L'étude des variables proportion de ... par page indique une différence significative seulement pour la variable

P3 (syntaxe), et encore là, c'est pour le groupe du troisième niveau que la moyenne est la plus basse: 0.127 par rapport à 0.216 pour le deuxième niveau et 0.194 pour le premier niveau.

D'autre part, si nous comparons le groupe des étudiants du premier niveau avec le groupe comprenant à la fois les étudiants de deuxième niveau et ceux de troisième niveau, et que nous effectuons les tests T pour vérifier l'hypothèse d'égalité entre les moyennes, nous devons encore rejeter cette hypothèse pour le nombre total d'erreurs par page (avec un niveau de confiance de 0.05). La moyenne est de 2.6662 pour le niveau 1 et de 2.1394 pour les deux autres. De plus, les valeurs du T sont assez élevées pour la grammaire et les anglicismes; bien que le test ne soit pas significatif pour ces deux catégories, c'est encore le groupe du niveau 1 qui affiche les plus hautes moyennes.

Nous pouvons donc conclure que le niveau d'études a une incidence directe sur le nombre des erreurs considérées globalement, comme le démontrent les tests F et T, et sur le nombre d'erreurs dans deux catégories: syntaxe et grammaire. Même là où les différences ne sont pas statistiquement significatives, la moyenne observée à un niveau inférieur est plus élevée que celle du niveau immédiatement supérieur, et ce pour toutes les catégories d'erreurs.

III

RELATIONS ENTRE LES VARIABLES

Nous nous proposons d'étudier dans cette section, les rapports entre les treize variables suivantes:

- NPC : nombre de pages composées
- MTVAR: nombre total de fautes par page
- MVAR1: nombre de fautes d'orthographe par page
- MVAR2: nombre de fautes de grammaire par page
- MVAR3: nombre de fautes de syntaxe par page
- MVAR4: nombre de fautes de sémantique par page
- MVAR5: nombre de renvois ambigüs par page
- MVAR6: nombre d'anglicismes par page
- P1 : proportion de fautes d'orthographe
- P2 : proportion de fautes de grammaire
- P3 : proportion de fautes de syntaxe
- P4 : proportion de fautes de sémantique
- P5 : proportion de renvois ambigüs
- P6 : proportion d'anglicismes

Pour tous les dossiers sans exception, nous possédions l'information nécessaire pour établir les valeurs de ces treize variables.

Rapport entre la variable NPC et les autres variables

Pour analyser les rapports entre le nombre de pages composées et les autres variables, nous avons d'abord divisé notre échantillon en deux groupes: ceux dont le texte a plus d'une page et ceux dont le texte a une page ou moins.

Voyons d'abord les variables P1 ... P6. Pour ce type de variable, il faut accepter l'hypothèse d'égalité dans tous les cas, que l'on admette que les deux groupes ont la même variance ou non. Les différences ne sont donc pas significatives.

Pour les sept autres variables, MTVAR et MVAR1 ... MVAR6, les résultats sont différents selon que l'on pose ou que l'on ne pose pas l'hypothèse que les deux groupes ont la même variance:

- Si on accepte cette hypothèse de variance égale:
Pour un niveau de confiance de 0.05, nous devons alors rejeter l'hypothèse d'égalité entre les moyennes des deux groupes pour les variables suivantes: MVAR1, MVAR2, MVAR6 et MTVAR. En d'autres termes, après avoir observé que le groupe qui a écrit moins d'une page obtient des moyennes plus élevées aussi bien dans le cas du nombre total de fautes que dans le cas des catégories orthographe, grammaire et anglicismes, il faut conclure que les différences sont statistiquement

significatives: le nombre de pages composées a une incidence sur le nombre global de fautes et sur le nombre de fautes de trois catégories différentes; et, contrairement à ce qu'on aurait eu tendance à prévoir, le nombre de fautes est alors grosso modo inversement proportionnel au nombre de pages composées, les étudiants ayant composé un texte plus long enregistrant les moyennes les plus basses.

- Si on rejette l'hypothèse de variance égale:

Dans un tel cas, il faut accepter l'hypothèse d'égalité entre les moyennes pour toutes les variables, sauf pour le nombre total de fautes (MTVAR); ajoutons que même pour cette dernière variable, la différence n'est significative que si l'on fixe le niveau de confiance à 0.05. Soulignons toutefois que les valeurs absolues du T fournies par les tests sont relativement élevées pour les catégories orthographe et grammaire; ce n'est là qu'un indice, mais un indice qui confirme notre précédente conclusion, à savoir que le nombre de fautes est grosso modo inversement proportionnel au nombre de pages composées. On peut d'ailleurs vérifier la validité de cette conclusion d'une autre façon: en effet, le groupe qui a composé moins d'une page obtient toujours des moyennes supérieures à celles de l'autre groupe pour les sept variables MTVAR et MVAR1 ... MVAR6, sauf dans le cas de la sémantique (MVAR4), comme le montre le tableau 20.

Tableau 20. Moyenne des deux groupes pour les variables MTVAR et MVAR1 ... MVAR6.

Variables	Groupe 1 Plus d'une page	Groupe 2 1 page et moins
MVAR1 Orthographe	0.7825	1.2824
MVAR2 Grammaire	0.8412	1.2430
MVAR3 Syntaxe	0.3832	0.4967
MVAR4 Sémantique	0.1522	0.0921
MVAR5 Renvois ambigüs	0.0248	0.0428
MVAR6 Anglicismes	0.0806	0.1974
MTVAR Total	2.2645	3.3543

Paires de variables

Comparons maintenant les moyennes des six variables MVAR1 ... MVAR6; pour toutes les paires de variables considérées, les tests T révèlent des différences significatives sauf pour la paire MVAR1 - MVAR2: en effet, la moyenne générale du nombre de fautes d'orthographe (MVAR1) par page est de 0.852 alors qu'elle est de 0.897 pour les fautes de grammaire (MVAR2); comme on le voit, si les moyennes ont tendance à prendre les mêmes valeurs dans ce dernier cas, cela n'est vrai pour aucune des autres paires de variables étudiées.

Nous avons obtenu des coefficients de corrélation pour toutes les paires de variables possibles. Dans les cas où

les coefficients de corrélation font intervenir l'une des variables P (P1 ... P6), les résultats ne nous permettent pas de tirer de conclusion.

Il n'en est pas de même lorsque les coefficients de corrélation font intervenir soit la variable MTVAR, soit l'une des six variables MVAR1 ... MVAR6. Ainsi, la variable MVAR6 (anglicismes) est corrélée positivement avec toutes les autres variables, sauf pour MVAR4 (sémantique), et de façon assez significative; cela veut dire qu'il y a parallélisme entre, d'une part le nombre d'anglicismes et d'autre part le nombre de fautes d'orthographe, de grammaire, de syntaxe et le nombre de renvois ambigüs: le nombre de fautes par page pour chacune de ces catégories a tendance à augmenter dans la même proportion que le nombre d'anglicismes. C'est pour les paires suivantes: MTVAR - MVAR1, MTVAR - MVAR2 et MTVAR - MVAR3 que nous enregistrons les coefficients de corrélation les plus élevés; dans tous les cas, ils sont positifs et significatifs: en d'autres termes, il existe un parallélisme certain entre d'une part le nombre global de fautes et d'autre part le nombre de fautes d'orthographe, de grammaire et de syntaxe. Rappelons que le calcul de coefficients de corrélation permet de démontrer que deux phénomènes quelconques vont de pair, mais non pas que l'un est la cause de l'autre.

De façon générale donc, la corrélation entre les divers types de fautes est assez forte, sauf dans le cas des renvois ambigüs et de la sémantique.

La deuxième partie de ce rapport a peut-être paru quelque peu abstraite et quelque peu difficile à certains lecteurs, et dans un tel cas, on peut penser que la terminologie technique employée y a sans doute été pour quelque chose. Mais puisqu'il s'agissait de statistique, on comprendra que nous ayions voulu nous en tenir à la terminologie de cette discipline et que c'était là la seule façon d'éviter l'équivoque, l'ambiguïté et la vulgarisation à outrance.

"Je joue avec les syllabes.
Je les frappe très fort, les unes contre les autres. Couvrir toutes les voix humaines qui pourraient monter et m'attaquer en foule. Dresser un fracas de syllabes rudes et sonores. M'en faire un bouclier de pierre. Une fronde élastique et dure. Kamouraska! Kamouraska! Il y a jonc au bord de l'eau!" (Anne Hébert)

TROISIÈME PARTIE

CONCLUSIONS
ET
RECOMMANDATIONS

Conclusions d'ordre statistique

Nous voulons faire maintenant la synthèse des résultats les plus pertinents de l'analyse statistique; il est toutefois difficile de résumer des faits sans donner l'impression de généraliser. Nous avons fait, au cours de l'analyse, toutes les nuances qui s'imposaient; c'est pourquoi nous demandons à nos lecteurs, au moment où ils prendront connaissance de la présente synthèse, de revoir le détail de l'analyse si nécessaire, afin d'éviter les généralisations abusives.

L'analyse statistique nous a permis de faire des découvertes qui ne manquent pas d'intérêt. Ainsi, on aurait pu s'attendre à ce que la moyenne cumulative ait une influence sur les variables, c'est-à-dire sur la qualité du français telle que nous l'avons définie; or, il n'en est rien ou presque, comme l'ont démontré les tests de vérification (pp.62-65). De même, une batterie de vingt-sept tests a permis de constater que le fait d'avoir fréquenté une institution préuniversitaire privée plutôt que publique n'influence pas de façon significative la qualité du français écrit des individus de l'échantillon (pp.61-62).

Un autre résultat – et non le moindre – c'est que les données statistiques que nous avons obtenues permettent de contester une impression fort répandue, à savoir que le français de nos étudiants serait rempli d'anglicismes et de fautes

de syntaxe (pp.47-48). D'autre part, était-il prévisible que le sexe (pp.56-58) et le lieu d'origine (pp.58-59) pouvaient avoir une influence si déterminante sur plusieurs variables?

Nous avons également établi que l'appartenance à tel niveau (pp.65-66) avait une influence certaine sur plusieurs variables, notamment sur le nombre global de fautes et sur le nombre de fautes de grammaire et de syntaxe. De la même façon, les résultats indiquent que le fait de poursuivre tel programme d'études (pp.53-56) plutôt que tel autre a une incidence sur les variables suivantes: nombre global de fautes, nombre de fautes de grammaire et d'orthographe. En d'autres termes, la qualité du français s'améliore, du moins pour certains aspects, à mesure que l'étudiant progresse dans ses études, et cette qualité diffère selon que le programme d'études est axé sur la langue ou non. Soulignons encore que si la qualité du français a tendance à changer de façon positive à mesure que l'étudiant atteint des niveaux supérieurs, ce changement n'est toutefois pas influencé par la progression en âge (pp.59-61). Nous avons pourtant remarqué que les étudiants les plus âgés (31 ans et plus) commettent globalement moins d'erreurs, font moins de fautes d'orthographe et de grammaire et rédigent des textes contenant moins d'anglicismes.

Quant à ceux qui auraient pu croire que des textes courts n'avaient aucune valeur pour l'enquête, il n'ont qu'à

revoir les résultats (pp. 68-70) pour constater que les auteurs peu prolifiques ont fait en moyenne plus de fautes que les autres et que, de plus, les différences sont statistiquement significatives pour trois variables, dont le nombre global de fautes.

Enfin, comme l'a montré l'analyse statistique, la corrélation entre la plupart des types d'erreurs est assez forte (pp.70-71), notamment entre les anglicismes et les autres types (sauf pour la sémantique) et entre le nombre global de fautes d'une part et les fautes de grammaire, d'orthographe et de syntaxe d'autre part. L'existence de cette corrélation, c'est-à-dire de la croissance parallèle de plusieurs types de fautes, aura une incidence sur les recommandations que nous pourrons faire, comme on le verra par la suite.

Conclusions d'ordre linguistique

Les étudiants de notre échantillon écrivent-ils un français qui mérite le qualificatif de "bon" ou celui de "mauvais"? Ainsi libellée, cette question que tous voudraient nous poser—certains l'ont déjà fait—ne peut que rester sans réponse; car nous avons le droit d'exiger, avant de répondre, qu'on nous propose une définition opératoire de ce qu'on entend par "bon" ou "mauvais" français —qualificatifs qui ne sauraient avoir cours dans une analyse strictement linguistique. Les étudiants, rappelons-le, ont fait en moyenne entre deux et trois fautes par page. Est-ce énhaurme? Est-ce terrible? Est-ce trop? Nous n'en savons rien. Non pas que nous refusions

de nous prononcer, mais donner notre opinion sans avoir de point de comparaison, c'est-à-dire de résultats qui proviendraient d'une autre enquête, ce serait émettre un simple jugement de valeur fondé sur le seul sentiment linguistique, ce qu'un linguiste doit éviter de faire. Or, ce point de comparaison n'existe pas; se prononcer, dans un tel cas, ce serait justement verser dans la généralisation abusive. Le rapport que nous soumettons fournit, nous semble-t-il, assez d'indications pertinentes sur la qualité du français d'un groupe donné par rapport à un autre groupe pour qu'on s'en tienne à ces indications. Le reste est affaire de corridors et de salons!

Le problème est le suivant: comment peut-on émettre un jugement sur un état de langue par rapport à un "modèle" de belle (?) langue, parlée ou écrite? Comment? en effet, lorsqu'on sait par exemple, en ce qui concerne l'aspect phonétique, que la prononciation proposée comme étant la bonne par beaucoup de dictionnaires n'est pas fondée, la plupart du temps, sur une enquête linguistique exhaustive mais seulement sur des inventaires partiels et, par surcroît, limités à une partie seulement de la francophonie — sans compter les nombreux ouvrages ou dictionnaires qui se contentent de postuler la "bonne" prononciation à partir d'autres ouvrages ... Dans leur Dictionnaire de la prononciation française dans son usage réel, André Martinet et Henriette Walter présentent trois prononciations effectivement entendues pour des mots aussi fré-

quents que traîner et vêtir, cinq prononciations pour des mots aussi courants que cuillerée et augmentation, six prononciations pour péréquation et genèvrier, etc. et cela, pour la seule région parisienne! Ce que ces deux linguistes écrivent à propos de la prononciation peut être appliqué aux autres aspects de la langue:

> "Est-ce à dire que toutes les prononciations sont recommandables? Et d'abord qu'est-ce qu'une prononciation recommandable? On est tenté de la définir comme celle que pratiquent les personnes et les classes qui jouissent, dans la société, du plus haut prestige. Il resterait alors à déterminer quelles sont ces personnes et ces classes. Il n'est pas sûr d'ailleurs que l'emploi, parmi des sujets appartenant à des milieux modestes, de prononciations de prestige n'entraînerait pas la moquerie ou, du moins, quelques sourires. En d'autres termes, n'importe quelle forme linguistique peut, dans certaines circonstances, être "déplacée". Il y aurait donc autant de normes qu'il y a de milieux distincts, et ceux qui ont à vivre dans différents milieux devraient, dans ce cas, apprendre à changer de norme d'un instant à l'autre. C'est effectivement ce que nous constatons chez la plupart de nos contemporains: on ne parle pas en famille et entre amis comme on le fait en public. (p. 16).

De même, il est évident que les étudiants qui ont participé à l'enquête n'écrivent pas de la même façon selon qu'ils rédigent un texte libre, un travail à remettre à leur professeur ou un exposé à lire devant leurs confrères. A ce propos, il est permis de faire l'hypothèse que la proportion d'erreurs ressortissant à la sémantique (ambiguïtés, équivoques, emplois de termes impropres) serait plus élevée dans un corpus de travaux académiques que celle que nous avons observée dans le corpus

de textes libres de notre échantillon. Cela tient, entre autres choses, au fait que chaque discipline possède un bagage de termes spécialisés et que cet ensemble terminologique constitue, pour les étudiants, l'une des premières difficultés rencontrées dans l'apprentissage de la discipline en question.

Une expérience limitée à quelques cours nous a permis de constater par exemple que plusieurs étudiants arrivent difficilement à faire concrètement la distinction entre les termes suivants: hypothèse, postulat, principe, méthodologie, méthode, analyse, procédure, description, tâche et objectif — qui sont pourtant des termes fondamentaux pour toute recherche ou toute étude dans une discipline donnée.

Nous n'avons pas voulu utiliser un corpus de travaux académiques pour notre enquête, parce qu'il nous fallait obtenir l'homogénéité quant à la nature du texte à produire, pour que les résultats fussent valables: comment aurions-nous pu procéder à des comparaisons valides si nous avions eu comme corpus des études sur les glaciations ou sur l'histoire de la Justice au Québec, des descriptions phonétiques et phonologiques d'un parler ou encore des travaux dans les domaines de la philologie, de la traduction, de la littérature et de la sémilogie de la peinture?

Et nous pensons que ceux qui voudraient, pour leur propre gouverne, compléter notre étude, auraient avantage à

constituer un corpus de textes, tout simplement parce que lorsqu'on veut observer un état de langue, il faut l'observer au naturel, c'est-à-dire à partir d'énoncés réels effectivement produits. Loin de nous l'idée de décrier les tests: seulement il faut bien voir qu'ils ont une fonction très différente: celle de vérifier l'état des connaissances des répondants sur un sujet donné. Dans un texte, le candidat utilise la langue, et comme bon lui semble; devant un test, le même candidat réfléchit sur la langue, et de manière obligée, car sa réflexion est orientée par le choix même des questions qu'on lui impose. Ajoutons que dans ce domaine, les tests, dans la majorité des cas, proposent à l'étudiant des exemples fabriqués de toutes pièces pour les besoins de la cause, de sorte que le candidat est appelé à réfléchir sur une langue en quelque sorte artificielle. La tâche d'un étudiant qui passe un test portant sur la syntaxe, la grammaire, l'orthographe, la sémantique ou les anglicismes est moins d'utiliser la langue que de faire état de ses connaissances sur l'un de ces sujets. Nous croyons que de tels tests font partie des moyens pédagogiques nécessaires à l'enseignement d'une langue, mais qu'ils sont moins indiqués lorsqu'il s'agit d'étudier la langue telle qu'elle est utilisée spontanément par un groupe quelconque. Attirer l'attention d'un candidat sur un point précis de la langue, comme le font les questions d'un test, c'est influencer le candidat. Et c'est ce que nous avons voulu éviter en demandant aux étudiants de rédiger un texte sur un sujet de leur choix.

Recommandations

Nous hésitons beaucoup à faire des recommandations, d'abord à cause du point de vue que nous avons adopté au cours de cette enquête de même que dans le présent rapport, et ensuite parce que notre mandat n'est pas très clair à ce sujet. Mais nous savons bien que les membres de plusieurs Comités et Commissions de même que plusieurs lecteurs s'attendent à ce que nous prenions position, car c'est bien de cela qu'il s'agit. En faisant les suggestions qui suivent, nous pensons que le principe d'objectivité énoncé au début de ce rapport garde toute sa valeur.

1. Qu'on ne s'étonne pas que notre première suggestion en soit une de prudence. Nous ne voulons pas en effet que notre rapport serve de caution à des décisions fondées sur des généralisations abusives qu'on se permettrait de faire et qui n'apparaissent pas ici.

2. A notre avis, les Comités ou Commissions de l'Université officiellement concernés par ce rapport ne devraient pas songer à proposer la création de cours de rattrapage en syntaxe pour tel groupe, en grammaire pour tel autre groupe, en orthographe pour telle autre catégorie d'étudiants et ainsi de suite, même si les résultats de notre analyse statistique démontrent clairement en faveur ou au dépens de quel groupe s'établit la différence constatée pour tel

aspect du français écrit. On comprendra que cette suggestion est une déduction logique de notre analyse, si l'on se rappelle que la corrélation entre les divers types d'erreurs est relativement forte.

3. Si donc ces Comités ou Commissions décident —car c'est à eux que revient la décision— de procéder à la création de cours, il faudrait que le contenu de ces cours touche tous les aspects de la langue écrite, quel que soit le groupe d'étudiants considérés.

4. En supposant toujours qu'une telle décision soit prise, nous suggérons que la formule adoptée ne soit pas de type négatif; la formule de cours dits "de rattrapage" n'est pas la meilleure. Elle est même à déconseiller, dans la mesure où elle obligerait à procéder à une "reclassification" des étudiants (après les "reclassifications" que nous avons subies au Québec récemment, on voit dans quel sens il faut prendre ce mot!).

5. De même, si une formule quelconque est adoptée, il ne faudrait pas qu'elle soit sélective: en d'autres termes, si une Direction d'études ou un Département par exemple voyait la nécessité d'améliorer la qualité du français écrit de ses étudiants, il serait opportun, nous semble-t-il, que tous les étudiants inscrits au programme d'études concerné participent à l'activité projetée. Car il faut éviter

la ségrégation entre les groupes d'une même Direction d'études ou d'un même Département.

6. Même si nous ne pouvons dire que le français de nos étudiants est bon ou mauvais, le rapport que nous présentons nous permet de souhaiter une plus grande conformité au code du français écrit.

   Et s'il y a lieu d'accorder au français une place de choix dans cette Université qui compte parmi les grandes institutions francophones d'Amérique et qui est située, par surcroît, dans l'une des plus importantes villes francophones du Québec, nous pensons que c'est l'affaire de tous, étudiants, professeurs et administrateurs. Et l'une des façons d'accorder à la langue française une place de choix dans notre Faculté —sinon dans notre Université—, ce serait de pouvoir offrir, non pas un cours d'appoint à quelques groupes, mais un cours de haut niveau et de grande qualité qui s'adresserait à l'ensemble des étudiants, tout en tenant compte du programme d'études de chaque groupe. En faisons-nous une recommandation? ...Ce n'est pas si simple... et il faudrait en discuter davantage; c'est en tout cas une façon d'affirmer que les expériences isolées et sporadiques comme les cours de rattrapage nous semblent vouées à l'échec.

7. Et si l'on devait un jour adopter cette dernière formule, que le cours en question ne soit pas un cours hors-programme,

mais un cours intégré à chacun des programmes et qui compte dans le calcul de la moyenne cumulative ... car les étudiants viennent aussi à l'Université pour obtenir un diplôme et ils paient assez cher pour cela!

Nous voulons ou nous ne voulons pas vivre en français. Nous voulons ou nous ne voulons pas enseigner et administrer en utilisant le français. Les décisions que nous devons tous prendre à propos de notre langue maternelle dépendent de cette volonté. Qu'aucun nouveau cours ne soit créé immédiatement à la suite de ce rapport ne nous étonnerait nullement, étant donné la nature de nos suggestions. Cependant, qu'absolument rien ne soit fait, à quelque niveau que ce soit, celui des administrateurs, celui des professeurs ou celui des étudiants, ne laisserait pas d'en surprendre plus d'un. On mesurerait alors la force et la qualité ... de cette volonté dont nous avons parlé plus haut.

ENVOI

Du primaire jusqu'à l'Université –et il ne semble pas que la tradition soit sur le point de se perdre! – les étudiants québécois se sont fait dire d'abord: "Vous prononcez mal", ensuite, par le biais d'un glissement de sens entraînant une généralisation: "Vous parlez mal", enfin: "Vous écrivez mal" et en conclusion ... "C'est du joual!"

A force de juger le mode d'expression d'un groupe, on finit par tuer l'expression même chez ce groupe! Depuis fort longtemps, on entretient au Québec une véritable psychose de la langue: est-ce là le meilleur moyen d'aider les écoliers, les étudiants et les Québécois en général à être "maîtres chez eux" de leur langue?

Avant d'entreprendre quoi que ce soit, il faut détruire ce climat de psychose dont nous, administrateurs, "spécialistes" et professeurs du primaire à l'Université sommes peut-être plus responsables que les étudiants. Par la suite, les étudiants ayant cessé de se sentir exacerbés par le faux problème du "bon" et du "mauvais" français seront-ils mieux en mesure de comprendre que le succès dans leurs études dépend, pour une large part, de la maîtrise qu'ils ont de leur langue maternelle, quel que soit le programme d'études poursuivi. Soyons positifs! Et nous pensons ici à l'excellent exemple

que nous donne, non pas un linguiste, mais un spécialiste de géomorphologie, Louis-Edmond Hamelin, qui a su se tailler une réputation internationale dans son domaine tout en ayant le mérite de créer une terminologie précise, opératoire et française qui porte sur les phénomènes de glaciation et qui fait autorité.

Dans Blanche ou l'oubli, Aragon écrit cette phrase admirable: "Chaque fois qu'il y a une révolution, en principe, il faudrait changer la grammaire". On peut sans doute inverser les termes et dire: quand la grammaire change, c'est qu'il y a eu une révolution. La grammaire a peut-être changé au Québec; en tout cas, on n'a pas craint, depuis quelques années, d'affirmer une grammaire et un lexique qu'on avait accoutumé de cacher par le passé, —et ce n'est pas forcément une mauvaise chose. C'est peut-être qu'il y a eu une révolution dans les esprits et la société. Révolution qui semble s'achever pour en préparer une autre. Il est peut-être temps, pour chacun, de renverser la vapeur, et dans le sens qui convient au peuple d'ici.

Neuville, 19 février 1976

Liste des schémas
et
des tableaux

Schéma I. Variables et paramètres----------26

Tableau I. Répartition des individus d'après le niveau---30

Tableau 2. Répartition des individus d'après la Direction d'études----------30

Tableau 3. Répartition des individus d'après le sexe-----31

Tableau 4. Répartition des individus d'après le lieu d'origine----------31

Tableau 5. Répartition des individus d'après l'institution préuniversitaire fréquentée----------31

Tableau 6. Répartition des individus d'après le niveau et la Direction d'études----------37

Tableau 7. Répartition des individus d'après le sexe et la Direction d'études----------38

Tableau 8. Répartition des individus d'après la Direction d'études et l'institution préuniversitaire fréquentée ----------39

Tableau 9. Répartition des individus d'après le lieu d'origine et la Direction d'études----------40

Tableau 10. Répartition des individus d'après l'âge et l'institution préuniversitaire fréquentée--41

Tableau 11. Répartition des individus d'après le lieu d'origine et l'institution préuniversitaire fréquentée----------42

Tableau 12. Répartition des individus d'après le niveau et l'institution préuniversitaire fréquentée--43

Tableau 13. Répartition des individus d'après le niveau et l'âge----------44

Tableau 14.Répartition des individus d'après le niveau et le sexe----------45

Tableau 15.Nombre de pages composées (variable NPC)------46

Tableau 16.Nombre de cas de non-conformité pour chaque catégorie----------48

Tableau 17.Proportion de cas de chaque catégorie---------50

Tableau 18.Analyse de variance: Sexe-MVAR1 (orthographe)-58

Tableau 19.Analyse de variance: Sexe-MVAR2 (grammaire)---58

Tableau 20.Moyenne des deux groupes pour les variables MTVAR et MVAR1...MVAR6----------70

TABLE DES MATIERES

PREMIERE PARTIE

L'ENQUETE

I Introduction---------- 4

Origine de l'enquête: p. 4

II Quelques aspects théoriques---------- 5

Linguistique et purisme: p. 5 -- Définitions: p. 6

III Les modalités de l'enquête---------- 10

Test ou texte: p. 10 -- La fiche de renseignements: p. 12 --Le déroulement de l'enquête: p. 13 -- La participation à l'enquête: p. 15 -- L'examen des copies: p. 18

DEUXIEME PARTIE

ANALYSE STATISTIQUE

Introduction---------- 23

I L'échantillon. La répartition des individus---------- 27

Caractère de l'échantillon et limitations: p. 27 --Répartition des individus d'après l'un des paramètres: p. 28 --Répartition des individus d'après deux paramètres: p. 32 --Niveau et direction d'études: p. 32 -- Direction d'études et sexe: p. 33 -- Direction d'études et institution préuniversitaire fréquentée: p. 34. -- Direction d'études et lieu d'origine: p. 34 -- Age et institution préuniversitaire fréquentée: p. 35 -- Lieu d'origine et institution préuniversitaire fréquentée: p. 35 -- Niveau et institution fréquentée: p. 35 -- Niveau et âge: p. 35. Niveau et sexe: p. 36 -- Tableaux 6 à 14: p. 37-45

II Variables et paramètres---------- 46

Nombre de pages composées: p. 46 -- Proportions pour chaque catégorie: p. 49 -- Méthode d'analyse: p. 51 --Direction d'études: p. 53 -- Sexe: p. 56 -- Lieu d'origine: p. 58 -- Age: p. 59 -- Institution préuniversitaire fréquentée: p. 61 -- Moyenne cumulative: p. 62 -- Niveau: p. 65

III Relations entre les variables---------- 67

Rapport entre la variable NPC et les autres variables: p. 63 -- Paires de variables: p. 70

TROISIEME PARTIE

CONCLUSIONS ET RECOMMANDATIONS

Conclusions d'ordre statistique: p. 74 -- Conclusions d'ordre linguistique: p. 76 -- Recommandations: p. 81

ENVOI---------- 85

Liste des schémas et des tableaux---------- 87